bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

SOFTWARE

DeDaL: Cytoscape 3.0 app for producing and morphing data-driven and structure-driven network layouts

Urszula Czerwinska[1,2,3], Laurence Calzone[1,2,3], Emmanuel Barillot[1,2,3] and Andrei Zinovyev[1,2,3*]

*Correspondence: Andrei.Zinovyev@curie.fr
[1]Institut Curie, 26 rue d'Ulm, Paris, FR
Full list of author information is available at the end of the article

Abstract

Background: Visualization and analysis of molecular profiling data together with biological networks are able to provide new mechanistical insights into biological functions. Currently, high-throughput data are usually visualized on top of predefined network layouts which are not always adapted to a given data analysis task. We developed a Cytoscape app which allows to construct biological network layouts based on the data from molecular profiles imported as values of nodes attributes.

Results: DeDaL is a Cytoscape 3.0 app which uses linear and non-linear algorithms of dimension reduction to produce data-driven network layouts based on multidimensional data (typically gene expression). DeDaL implements several data pre-processing and layout post-processing steps such as continuous morphing between two arbitrary network layouts and aligning one network layout with respect to another one by rotating and mirroring. Combining these possibilities facilitates creating insightful network layouts representing both structural network features and the correlation patterns in multivariate data.

Conclusions: DeDaL is the first method allowing to construct biological network layouts from high-throughput data. DeDaL is freely available for downloading together with step-by-step tutorial at http://bioinfo-out.curie.fr/projects/dedal/.

Keywords: data dimension reduction; network layout; principal manifolds

Background

One of the major challenges in systems biology is to combine in a meaningful way the large corpus of knowledge in molecular biology recapitulated in the form of large interaction networks together with high-throughput omics data produced at increasing rate, in order to advance our understanding of biology or pathology [1].

There exists numerous methods using biological networks for making insightful high-throughput data analysis [1]. One can distinguish three large groups of such methods: (1) mapping the data on top of a pre-defined biological network layout, (2) identification of subnetworks in a global network possessing certain properties computed from the data (such as subnetworks enriched with differentially expressed genes), (3) using biological network structure for pre-processing the high-throughput data (for example, for "smoothing" the discrete mutation data).

Quantitative omics data can be mapped on top of a pre-defined biological network layout. Currently, most of the pathway databases (such as KEGG [2], Reactome

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

[3]) provides such a possibility, using simple data visualization tools. Omics data visualization tools using networks are constantly improved and become more elaborated [4]. For example, VANTED tool [5] creates a classification tree according to the KEGG pathway hierarchy and shows a biological network with omics data as barplots or pie-charts attached to the nodes which allows to visualize more complex data than by simple node coloring. NaviCell [6] and related pathway database Atlas of Cancer Signalling Network (ACSN) together with standard heat maps and barplots provide more flexible data visualization tools such as glyphs (symbols with configurable shape, size and color) and map staining (using the network background for visualization) [7]. An interesting approach for data visualization using biological networks was developed in NetGestalt online tool [8] which uses a NetSAM R package to create modules by hierarchical ordering of the network in one dimension and visualizes high-throughput data accordingly to a chosen track as a combination of barplots and heat maps.

Omics data are used to identify overexpressed or enriched subnetworks. For example, in [9] expression data were combined with network information in order to identify under- or overexpressed subnetworks in Huntington‘s disease and breast cancer. Inspired by this method, several Cytoscape plug-ins were developed and applied to various omics data in order to find connected sub-components where most of the genes are differentially expressed or co-expressed [10, 11]. A recent review on integrating molecular profiles with networks in order to find “network modules” can be found in [12].

Using projection of the high-throughput data into the basis of functions smooth on a biological network graph was suggested in [13]. Recently, biological networks were used to regularize the genome-wide mutational landscapes (which are sparse) in cancer, using network smoothing methods [14].

However, none of the methods cited above had a purpose to visualize high-throughput data by computing a specific network layout based on the omics data themselves, which would combine both network structure and the data from the network node attributes. Some of the existing Cytoscape layout algorithms (such as Group Attributes Layout) allow using the values of single node attributes, but this possibility is currently under-developed. We believe that using networks for visualizing and analyzing data requires methods that would be able to create more suitable and adapted for a particular task biological network layouts.

Mathematically speaking, molecular entities exist in two metric spaces: in space of biological functions, where the distance between two molecules can be defined by the number of steps (edges) in a graph defining pairwise functional relations (such as protein-protein interactions) along the shortest path connecting them; and in data space, where the distance between two molecules is defined by the proximity of the corresponding numerical descriptors (such as expression profiles). The network distances are usually visualized by designing a 2D or 3D layout, representing the network structure. Visualization of distances in data space is achieved by data dimension reduction methods (such as PCA) projecting multidimensional vectors in 2D or 3D space. No methods were developed so far for performing dimension reduction in Cytoscape and mixing the two types of visual representations together. Having this in mind, we’ve developed DeDaL, a Cytoscape 3.0 app for mixing purely data-driven and purely structure-driven network layouts.

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Implementation

DeDaL is a *simplified* Cytoscape 3 app implemented in Java language. For computing linear and non-linear principal manifolds, DeDaL uses VDAOEngine Java library, developed by AZ (http://bioinfo-out.curie.fr/projects/elmap/). For computing the eigenvectors of a symmetric Laplacian matrix, Colt library has been used (http://acs.lbl.gov/ACSSoftware/colt/). Internal graph implementation is re-used from BiNoM Cytoscape plugin [15, 16, 17]. The source code of DeDaL is available at http://bioinfo-out.curie.fr/projects/dedal.

Producing data-driven network layouts

Data-driven network layout (DDL) is produced by DeDaL by positioning the nodes of the network according to their projection from the multidimensional data space of associated numerical vectors into some 2D space. DeDaL implements three algorithms for performing this dimension reduction: (1) Projection onto a plane of two selected principal components; (2) Projection onto a non-linear 2D surface approximating the multidimensional data distribution, i.e. principal manifold, computed by the method of elastic maps [18, 19, 20, 21]; (3) Using (1) or (2) preceded by network-based regularization (smoothing) of the data, based on computing the k first eigen vectors of the Laplacian matrix of the network graph and projecting data into this subspace (as suggested in [13]).

DeDaL implements specific data pre-processing and resulting layout post-processing steps. Pre-processing steps include (1) selecting only nodes whose associated numerical vectors (imported as tables to Cytoscape) are sufficiently complete and (2) optional double centering of the data matrix. Post-processing of the resulting layout includes (1) avoiding overlap between node positions by moving them in a random direction at a small distance; (2) moving the outliers (nodes positioned too distantly from other nodes) closer to the barycenter of the data distribution; (3) placing the nodes with missing data into the mean point of the position of their network neighbours.

In the future we will exploit a possibility to project the data into 3D and will implement additional dimension reduction algorithms such as multidimensional scaling.

Manipulating network layouts in DeDaL

In order to allow comparing the resulting DDLs with standard layouts produced by Cytoscape and transform one into another, DeDaL implements simple layout morphing and aligning methods. Morphing of two network layouts is performed by a linear transformation, moving matched nodes along straight lines. DeDaL provides a convenient user dialog for morphing one layout into another in which a user can use slider and immediately appreciate the morphing result. The morphing operation provides poor results if one layout is systematically rotated or flipped with respect to the node positions in another one. DeDaL allows aligning two network layouts by rotating and mirroring, and minimizing the Euclidean distance between two layouts.

Double-centering the data matrix

The data matrix is optionally double-centered by subtracting from each matrix entry the mean value calculated over the corresponding matrix row and the mean

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

value calculated over the matrix column, and by adding the global mean value computed over all matrix entries. This procedure allows to eliminate some global biases in the data such as the global differences in average fluorescence intensity of different probes in microarray data.

Network-based smoothing of data

Network data smoothing is made in DeDaL as it was suggested in [13]. For a graph representing the biological network, its Laplacian and all its eigenvectors are computed. These vectors define a new orthonormal basis in the multidimensional data space. To smooth the values of the data matrix, the initial multidimensional vector associated to a datapoint is projected into the subspace spanned by the first k eigenvectors of the graph's Laplacian. DeDaL smoothing parameter is the $p_{ns} = 1 - \frac{k-(n_c+2)}{N-(n_c+2)}, p_{ns} \in [0;1]$, where n_c is the number of connected components in the graph and N is the number of nodes on the graph. Therefore, $p_{ns} = 0$ corresponds to $k = N$, i.e. when no smoothing is performed and all eigenvectors are used, while $p_{ns} = 1$ corresponds to $k = (n_c+2)$ and first two non-degenerated eigenvectors are used to smooth the data (the data become effectively three-dimensional, with the first dimension corresponding to the average value of the data matrix computed over each connected component of the graph).

Exporting the pre-processed data

The results of pre-processing the data for a given network can be exported to a file. Actually, two files are created: one in a simple tab-delimited format suitable for further analysis in most statistical software packages and another file in the ".dat" format, suitable for analysis in ViDaExpert multidimensional data visualization tool [22]. This possibility can be used, for example, for network smoothing of an expression dataset for further application in any machine learning algorithm (clustering, classification). For this purpose, DeDaL can be also used in a command line mode (see examples at the web-site).

Computing principal components

The principal components in DeDaL are computed using singular value decomposition, computed by the method allowing to use missing data values without pre-imputing them, as it is described in [23]. Data points, containing more than 20% of missing values are filtered out from the analysis. DeDaL computes the 10 first principal components if there is more than 10 data points, and k principal components if there is $k + 1$ data points, $k < 10$. After computing the principal components, DeDaL reports the amount of variance explained by each of the principal components.

Continuous layout morphing

Morphing two network layouts is performed by a simple linear transformation. A node having position (x_{11}, x_{12}) in the initial layout and the position (x_{21}, x_{22}) in the target layout is placed during the morphing procedure in the position $(p \times x_{21} + (1-p)x_{11}, p \times x_{22} + (1-p)x_{12})$, where $p \in [0;1]$ is the morphing parameter representing the fracture of distance between the initial and target node positions along the straight line.

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

Aligning two network layouts by rotation and mirroring

Morphing between two network layouts might be meaningless if all nodes in one layout are systematically rotated or flipped with respect to the node positions in another layout. This situation is often the case when producing the pure data-driven layout and comparing it to the initial structure-driven layout. In this case, DeDaL allows minimizing the Euclidean distance between two layouts defined as the sum of squared Euclidean distances between all matched nodes with respect to all possible rotations and mirroring of one of the layouts. To do this, a user should simply check the corresponding checkbox in the user dialog before starting to apply layout morphing. Also, a user can align several network layouts to one chosen reference network layout, using a separate "Layout aligning" dialog. For example, it is usually useful to align the structure-driven layouts to the PCA-based data-driven layout.

Results

Using The Cancer Genome Atlas transcriptome data and Human Protein Reference Database network

We used The Cancer Genome Atlas (TCGA) transcriptomic dataset for breast cancer (548 patients)[24] and Human Reference Protein Database (HRPD) database [25] as a source of protein-protein interaction network.

Firstly, as an example of a small subnetwork, we selected proteins involved in Fanconi DNA repair pathway [26] as it is defined in Atlas of Cancer Signaling Network (ACSN, `http://acsn.curie.fr`). For node coloring, we mapped the value of the t-test computed for the gene expression difference between the basal-like (one of the molecular subtypes of breast cancer, significantly contributing to the intertumoral variability) and non basal-like breast tumours. We've imported the TCGA data in Cytoscape and applied DeDaL for the transcription levels of the genes in the subnetwork (Figure 1).

One can see (Figure 1, top right) that the first principal component sorts the nodes accordingly to the t-test, because in this case the first principal component is associated with the basal-like breast cancer subtype. The second principal component gives additional information such as that the expression levels of BRCA2 and FANCE are differently modulated though both are upregulated in the basal-like subtype. Morphing the organic network layout with the PCA-based layout moves position of some of the genes, keeping the general pattern of PCA preserved, while better reflecting the network structure.

We've also applied PCA-based DDL to the subset of basal-like breast tumours (Figure 1, bottom left) which showed the specific role of BRCA1 gene in this subtype (which is known). Also, the position of USP1 gene has significantly changed with respect to the PCA-based DDL produced for the whole set of samples. This demonstrates the ability of DeDaL to produce network layouts specific for a particular cancer subtype.

Application of network smoothing is demonstrated at Figure 1, bottom middle. The layout preserves the general pattern of the PCA-based DDL, while better visualizing the network structure, and moving some proteins into a different position. For example, BRCA1 gene is moved to left because it is connected to several genes overexpressed in basal-like breast cancer subtype. Figure 1, bottom right, shows

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

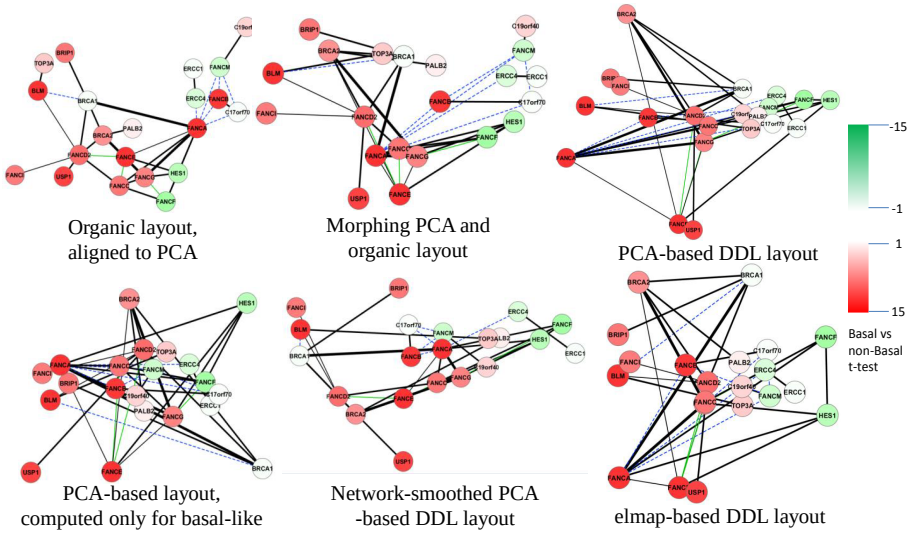

Figure 1 Using DeDaL for visualizing Fanconi pathway in breast cancer Top row from left to right: Standard organic layout, PCA-based DDL, morphing two previous layouts at half-distance. Bottom row from left to right: PCA-based DDL computed only for basal-like tumours (note change in position of BRCA1 gene), PCA applied to network-smoothed profile, DDL computed using elastic map (elmap) algorithm for computing non-linear principal manifold.

application of non-linear PCA to data dimension reduction. This network layout better resolves the relations between some gene expression levels such as FANCF and HES1 and the roles of BRCA1 and BRCA2 in Fanconi DNA repair pathway.

DDLs produced by DeDaL can serve to better visualize expression pattern in individual samples. Examples of using elastic map (elmap)-based DDL for distinguishing one randomly chosen basal-like and one non basal-like expression profiles of Fanconi pathway is shown in Figure 2. Unlike organic layout, DDL allows quickly evaluate the general trend of the expression profile and detect exceptions from this trend like USP1 gene, known to be a biomarker of genomic instability and Fanconi anemia phenotype [27], and overexpressed in both samples.

Secondly, we selected all proteins interacting with ESR1 protein (Figure 3). In this case, the second principal component shows, for example, that the expression levels of EGFR and CCNE1 are differently modulated though both are upregulated in the basal-like subtype. PCA layout also highlights a particular pattern of expression of some hub genes such as AR or EGFR, and shows that underexpressed genes in basal-like subtype forms more tightly connected subnetwork. Morphing the original organic network layout with the PCA-based layout moves position of some of the proteins, keeping the general pattern of PCA preserved. For example, underexpressed PIK3R1, IGFR1 and ERBB2 genes are moved on the left because each of them is connected to several overexpressed genes. Application of network smoothing drives the hub genes to the center of the layout, because of averaging over the hub's neighbors. It produces more regular pattern of network connections but approximately conserves the neighborhood relations in PCA layout. Therefore, combining DeDaL methods allows different ways of mixing network structure and high-throughput data for producing new network layouts.

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

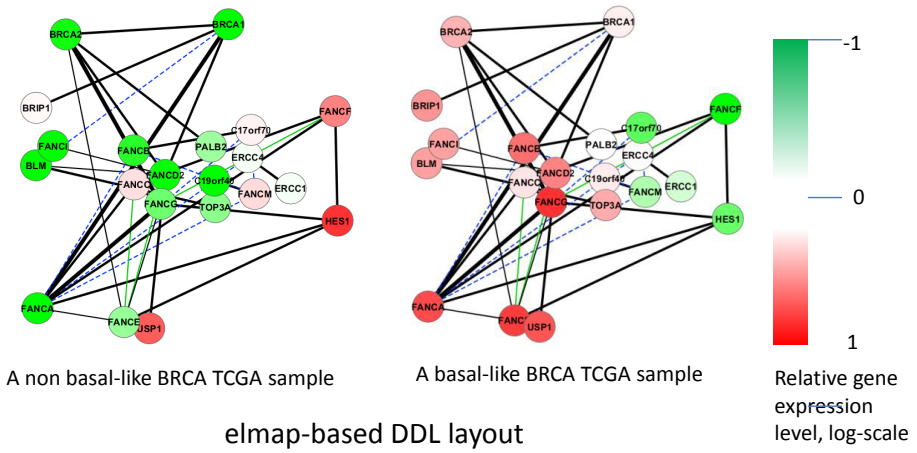

Figure 2 Using DeDaL for showing individual sample gene expression profiles. Expression profiles on the Fanconi pathway genes for two randomly chosen samples (one basal-like and one non basal-like) from TCGA breast cancer cohorts are shown. The expression levels are computed as relative to the mean value over the whole cohort.

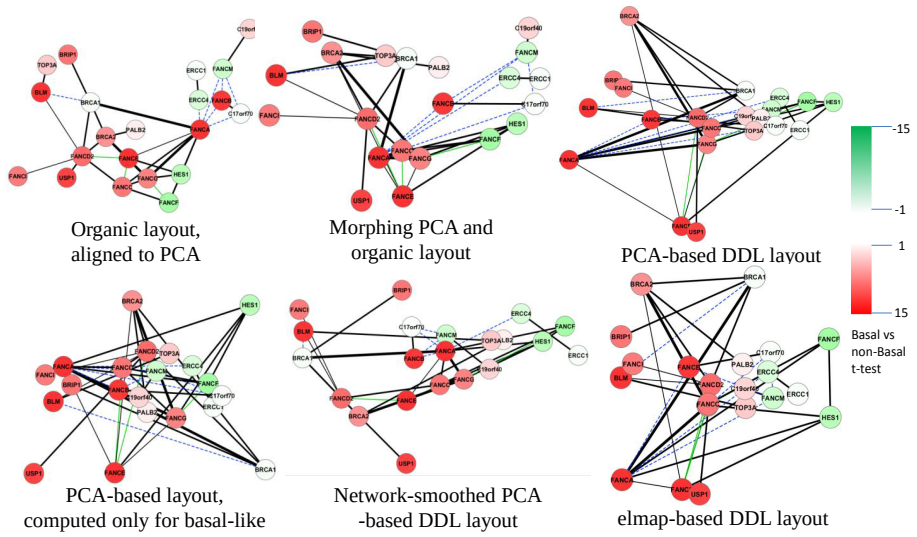

Figure 3 Using DeDaL for visualizing network of genes interacting with ESR1. DeDaL allows mixing purely structure-driven network layout (top left) with purely data-driven network layout (top right) by morphing them (bottom left, which is the half-distance between two upper layouts). Bottom right is the same as PCA-based layout (top right) but network smoothing was performed before applying PCA.

Visualizing genetic interactions

Genetic interactions between two genes happen in the case where their functions are synergistic (negative interactions) or mutually alleviating (positive interactions). The strength of genetic interactions is characterized by an epistatic score which quantifies deviation from a simple multiplicative model. In the global network of genetic interactions, each gene can be characterized by its epistatic profile, which is a vector of epistatic scores with all other genes [28]. It is shown that the genes with similar epistatic profiles tend to have similar cellular functions.

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

We applied DeDaL to create a DDL layout for a group of yeast genes involved in DNA repair and replication. The genetic interactions between these genes and the epistatic profiles (computed only with respect to this group of genes) were used from [28]. The definitions of DNA repair pathways were taken from KEGG database [2]. Figure 4 shows the difference between application of the standard organic layout for this small network of genetic interactions and PCA-based DDL (computed here without applying data matrix double-centering to take into account tendencies of genes to interact with smaller or larger number of other genes). PCA-based DDL in this case groups the genes with respect to their epistatic profiles. Firstly, local hub genes RAD27 and POL32 have distinct position in this layout. Secondly, PCA-based DDL roughly groups the genes accordingly to the DNA repair pathway in which they are involved. For example, it shows that Non-homologous end joining DNA repair pathway is closer to Homologous recombination (HR) pathway than to the Mismatch repair pathway. It also underlines that some homologous recombination genes (such as RDH54) are characterized by a different pattern of genetic interactions than the "core" HR genes RAD51, RAD52, RAD54, RAD55,

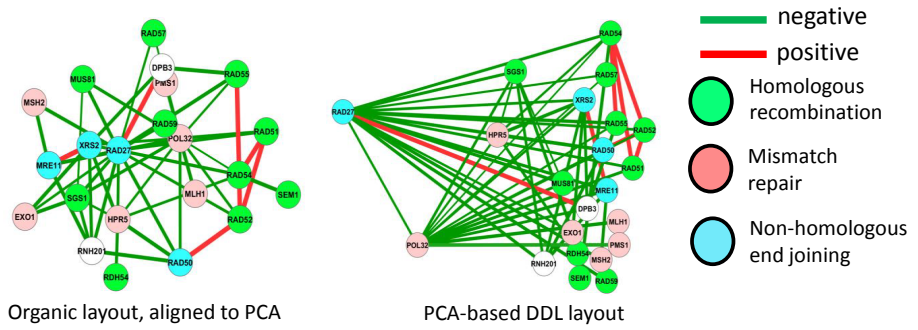

Figure 4 Using DeDaL for visualizing network of genetic interactions between yeast genes involved in DNA repair. Red and green edges denote positive and negative genetic interactions correspondingly. Different node colors indicate three distinct DNA repair pathways in yeast.

Visualizing attractors of a Boolean model

In this example we used the Boolean model of cell fate decisions between survival, apoptosis and non-apoptotic cell death (such as necrosis) published in [29], to group the nodes of the influence diagram accordingly to their co-activation patterns in the logical steady states. The table of steady states was taken from [29] (Figure 5, top right) and used to compute the PCA-based DDL (Figure 5, bottom left). In this DDL, nodes in close positions have similar pattern of activation in steady states (such as RIP1 and RIP1K). We used morphing PCA-based DDL and the initial layout of the model (as it was designed in [29]) to visualize several stable states corresponding to different cell fates (Figure 6). In this layout co-activated nodes tend to form compact groups. Therefore, DeDaL can be used to design layouts of mathematical models of biological networks, using the solutions of the model.

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

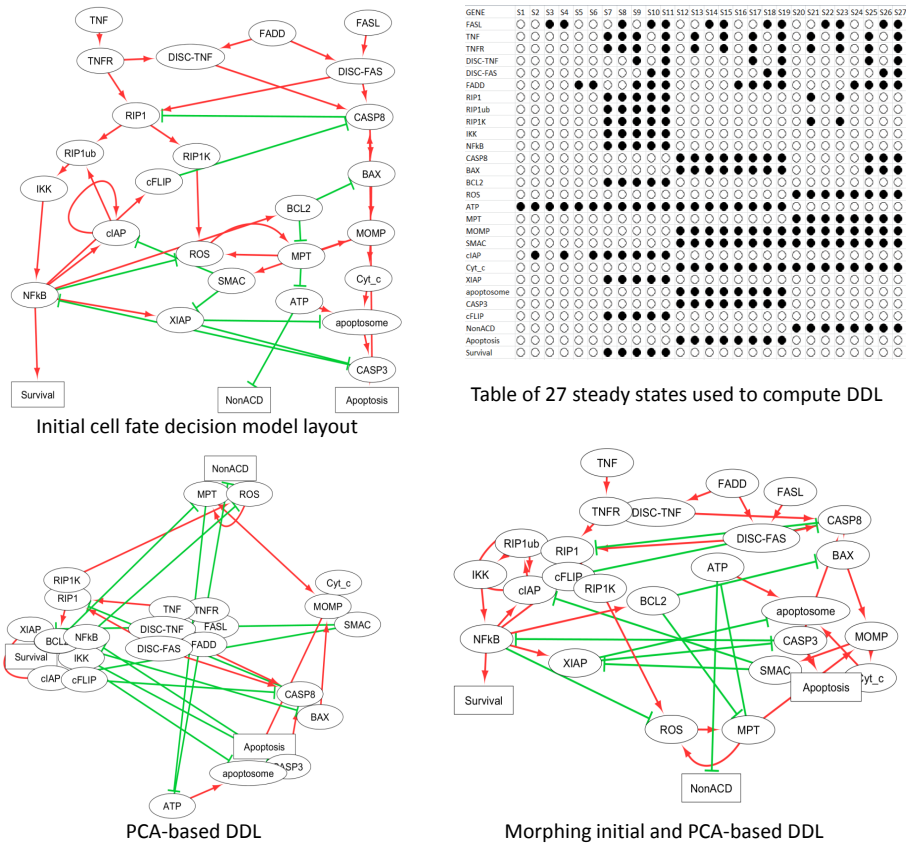

Figure 5 Using DeDaL for visualizing results of a Boolean model simulation. Table of computed steady states is used to group the nodes with similar states in similar conditions (shown in top right corner). In the influence diagram green edges signify inhibitory and red edges - activating relations.

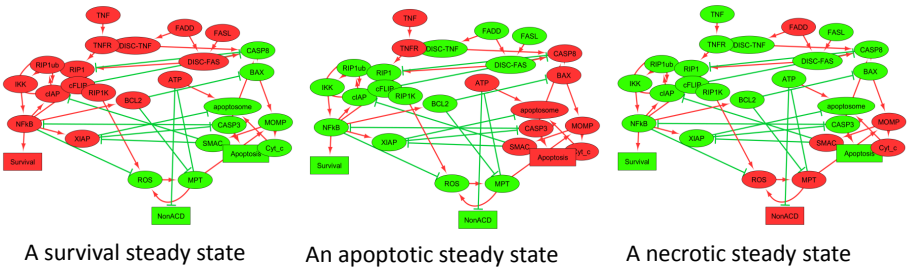

Figure 6 Using DeDaL for visualizing results of a Boolean model simulation. Visualization of three steady states of the model, with green and red denoting inactive (FALSE) and active (TRUE) states of the node correspondingly.

Conclusions

DeDaL Cytoscape plugin combines the classical and advanced data dimension reduction methods with the algorithms of network layouting inside Cytoscape environment. This ability can be used in a number of ways and applications, some of them are suggested in this manuscript.

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

The application of DeDaL is not limited to producing data-driven network layouts. More generally, DeDaL allows applying dimension reduction of the multivariate data associated with the nodes of any Cytoscape network, optionally using the structure of the network, and export the results for further analysis by any suitable algorithms.

In the future works an effort will be made to project the data in the three dimensional space. The user will be able to rotate freely in all three dimensions and better see patterns which are difficult to represent in 2D space. The software will be also completed with alternative dimension reduction algorithms such as multidimensional scaling which will extend the data modeling possibilities, better answering to specific user's needs.

Availability and requirements

Project name: DeDaL: Data-Driven Network Layouting

Project home page: http://bioinfo-out.curie.fr/projects/dedal/

Operating system(s): Platform independent

Programming language: Java

Other requirements: Java 1.6 or higher, Cytoscape 3.0 or higher

License: GNU GPL

Any restrictions to use by non-academics: free for any non-commercial use

Competing interests

The authors declare that they have no competing interests.

Author's contributions

AZ and UC developed the algorithm. UC and AZ wrote the application code. UC, LC and AZ provided the examples of using DeDaL. AZ and UC wrote the manuscript. AZ, LC, EB have read and revised the manuscript.

Acknowledgements

We acknowledge Eric Viara and Eric Bonnet for their help in implementing DeDaL and Loredana Martignetti for helping analysing the data. All authors are members of the team "Computational Systems Biology of Cancer". The work is supported by ITMO Cancer SysBio program, (INVADE project) and, the grant "Projet Incitatif et Collaboratif: Computational Systems Biology Approach for Cancer" from Institut Curie and by Institut National de la Santé et de la Recherche Médicale (U900 budget).

Author details

[1]Institut Curie, 26 rue d'Ulm, Paris, FR. [2]INSERM U900, Paris, FR. [3]Mines Paris Tech, Fontainebleau, FR.

References

1. Barillot, E., Calzone, L., Hupe, P., Vert, J.-P., Zinovyev, A.: Computational Systems Biology of Cancer. Chapman & Hall, CRC Mathemtical and Computational Biology, ??? (2012)
2. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., Tanabe, M.: Kegg for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res **40**(Database issue), 109–114 (2012). doi:10.1093/nar/gkr988
3. Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., Jupe, S., Kalatskaya, I., Mahajan, S., May, B., Ndegwa, N., Schmidt, E., Shamovsky, V., Yung, C., Birney, E., Hermjakob, H., D'Eustachio, P., Stein, L.: Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res **39**(Database issue), 691–697 (2011). doi:10.1093/nar/gkq1018
4. Gehlenborg, N., O'Donoghue, S.I., Baliga, N.S., Goesmann, A., Hibbs, M.A., Kitano, H., Kohlbacher, O., Neuweger, H., Schneider, R., Tenenbaum, D., Gavin, A.-C.: Visualization of omics data for systems biology. Nat Methods **7**(3 Suppl), 56–68 (2010). doi:10.1038/nmeth.1436
5. Klukas, C., Schreiber, F.: Integration of -omics data and networks for biomedical research with vanted. J Integr Bioinform **7**(2), 112 (2010). doi:10.2390/biecoll-jib-2010-112
6. Kuperstein, I., Cohen, D.P.A., Pook, S., Viara, E., Calzone, L., Barillot, E., Zinovyev, A.: Navicell: a web-based environment for navigation, curation and maintenance of large molecular interaction maps. BMC Syst Biol **7**, 100 (2013). doi:10.1186/1752-0509-7-100
7. Kuperstein, I., Grieco, L., Cohen, D., Thieffry, D., Zinovyev, A., Barillot, E.: The shortest path is not the one you know: application of biological network resources in precision oncology research. Mutagenesis **in press** (2015)
8. Shi, Z., Wang, J., Zhang, B.: Netgestalt: integrating multidimensional omics data over biological networks. Nat Methods **10**(7), 597–598 (2013). doi:10.1038/nmeth.2517

bioRxiv preprint doi: https://doi.org/10.1101/014365; this version posted January 26, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC 4.0 International license.

9. Ulitsky, I., Shamir, R.: Identification of functional modules using network topology and high-throughput data. BMC Syst Biol **1**, 8 (2007). doi:10.1186/1752-0509-1-8
10. Cline, M.S., Smoot, M., Cerami, E., et al, K.: Integration of biological networks and gene expression data using cytoscape. Nat Protoc **2**(10), 2366–2382 (2007)
11. Alcaraz, N., Friedrich, T., Kötzing, T., Krohmer, A., Müller, J., Pauling, J., Baumbach, J.: Efficient key pathway mining: combining networks and omics data. Integr Biol (Camb) **4**(7), 756–764 (2012)
12. Mitra, K., Carvunis, A.-R., Ramesh, S.K., Ideker, T.: Integrative approaches for finding modular structure in biological networks. Nat Rev Genet **14**(10), 719–732 (2013). doi:10.1038/nrg3552
13. Rapaport, F., Zinovyev, A., Dutreix, M., Barillot, E., Vert, J.-P.: Classification of microarray data using gene networks. BMC Bioinformatics **8**, 35 (2007). doi:10.1186/1471-2105-8-35
14. Hofree, M., Shen, J.P., Carter, H., Gross, A., Ideker, T.: Network-based stratification of tumor mutations. Nat Methods **10**(11), 1108–1115 (2013). doi:10.1038/nmeth.2651
15. Zinovyev, A., Viara, E., Calzone, L., Barillot, E.: Binom: a cytoscape plugin for manipulating and analyzing biological networks. Bioinformatics **24**(6), 876–877 (2008). doi:10.1093/bioinformatics/btm553
16. Bonnet, E., Calzone, L., Rovera, D., Stoll, G., Barillot, E., Zinovyev, A.: Practical use of binom: a biological network manager software. Methods Mol Biol **1021**, 127–146 (2013)
17. Bonnet, E., Calzone, L., Rovera, D., Stoll, G., Barillot, E., Zinovyev, A.: Binom 2.0, a cytoscape plugin for accessing and analyzing pathways using standard systems biology formats. BMC Syst Biol **7**, 18 (2013)
18. Gorban, A., Zinovyev, A.Y.: Method of elastic maps and its applications in data visualization and data modeling. International Journal of Computing Anticipatory Systems, CHAOS **12**, 353–369 (2001)
19. Gorban, A., Zinovyev, A.: Elastic principal graphs and manifolds and their practical applications. Computing **75**(4), 359–379 (2005)
20. Gorban, A., Kegl, B., Wunsch, D., Zinovyev, A. (eds.): Principal Manifolds for Data Visualisation and Dimension Reduction, LNCSE 58. Springer, ??? (2008)
21. Gorban, A.N., Zinovyev, A.: Principal manifolds and graphs in practice: from molecular biology to dynamical systems. Int J Neural Syst **20**(3), 219–232 (2010). doi:10.1142/S0129065710002383
22. Gorban, A.N., A., P., Zinovyev, A.: Vidaexpert: user-friendly tool for nonlinear visualization and analysis of multidimensional vectorial data. Arxiv preprint (1406.5550) (2014)
23. Gorban, A.N., Zinovyev, A.: Principal graphs and manifolds. In Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods and Techniques, eds. Olivas E.S., Guererro J.D.M., Sober M.M., Benedito J.R.M., Lopes A.J.S. (2009)
24. TCGA: Comprehensive molecular portraits of human breast tumours. Nature **490**(7418), 61–70 (2012). doi:10.1038/nature11412
25. Peri, S., Navarro, J.D., Kristiansen, T.Z., Amanchy, R., Surendranath, V., Muthusamy, B., Gandhi, T.K.B., Chandrika, K.N., Deshpande, N., Suresh, S., Rashmi, B.P., Shanker, K., Padma, N., Niranjan, V., Harsha, H.C., Talreja, N., Vrushabendra, B.M., Ramya, M.A., Yatish, A.J., Joy, M., Shivashankar, H.N., Kavitha, M.P., Menezes, M., Choudhury, D.R., Ghosh, N., Saravana, R., Chandran, S., Mohan, S., Jonnalagadda, C.K., Prasad, C.K., Kumar-Sinha, C., Deshpande, K.S., Pandey, A.: Human protein reference database as a discovery resource for proteomics. Nucleic Acids Res **32**(Database issue), 497–501 (2004). doi:10.1093/nar/gkh070
26. Moldovan, G.-L., D'Andrea, A.D.: How the fanconi anemia pathway guards the genome. Annu Rev Genet **43**, 223–249 (2009). doi:10.1146/annurev-genet-102108-134222
27. Kim, J.M., Parmar, K., Huang, M., Weinstock, D.M., Ruit, C.A., Kutok, J.L., D'Andrea, A.D.: Inactivation of murine usp1 results in genomic instability and a fanconi anemia phenotype. Dev Cell **16**(2), 314–320 (2009). doi:10.1016/j.devcel.2009.01.001
28. Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., Prinz, J., St Onge, R.P., VanderSluis, B., Makhnevych, T., Vizeacoumar, F.J., Alizadeh, S., Bahr, S., Brost, R.L., Chen, Y., Cokol, M., Deshpande, R., Li, Z., Lin, Z.-Y., Liang, W., Marback, M., Paw, J., San Luis, B.-J., Shuteriqi, E., Tong, A.H.Y., van Dyk, N., Wallace, I.M., Whitney, J.A., Weirauch, M.T., Zhong, G., Zhu, H., Houry, W.A., Brudno, M., Ragibizadeh, S., Papp, B., Pál, C., Roth, F.P., Giaever, G., Nislow, C., Troyanskaya, O.G., Bussey, H., Bader, G.D., Gingras, A.-C., Morris, Q.D., Kim, P.M., Kaiser, C.A., Myers, C.L., Andrews, B.J., Boone, C.: The genetic landscape of a cell. Science **327**(5964), 425–431 (2010). doi:10.1126/science.1180823
29. Calzone, L., Tournier, L., Fourquet, S., Thieffry, D., Zhivotovsky, B., Barillot, E., Zinovyev, A.: Mathematical modelling of cell-fate decision in response to death receptor engagement. PLoS Comput Biol **6**(3), 1000702 (2010)